

RunSemblance Documentation

Makes a semblance correlation of the two input traces. This is a full (non-gated) semblance.

Usage

The program is used as follows:

```
runSemblance input1FileName input2FileName
```

Where the `input1FileName` and `input2FileName` are in our standard [file format](#).

Output

The output of the program appears one place:

- (1) A copy of the output data is placed in the file `lastDataOutput`.

The returned floating point array is numbered $[0, N_f)$, and the i^{th} point represents the value of the semblance when the middle of the event is lined up with the i^{th} point.

Theory

We present here the definition of semblance implemented in the [LwsVector](#) class using the method `Semblance`. We start by reviewing the basic theory of cross correlation. Our implementation of cross correlation is defined in [RunXCorr](#). We summarize below the theory of cross correlation with lag.

Cross Correlation with Lag:

Define:

$$(1.1) \phi_{f,g}(t) = \int_{-\infty}^{\infty} f(\tau)g(t+\tau)d\tau$$

or in the digital domain:

$$(1.2) \phi_{f,g}(l,k) = \sum_{i=k-N/2}^{i=k+N/2} f(i)g(i+l)$$

where l is the lag, and k is the center of the window.

This should be normalized by

$$(1.3) \left[\sum_{i=k-N/2}^{i=k+N/2} f^2(i) \sum_{i=k-N/2}^{i=k+N/2} g^2(i+l) \right]^{1/2}$$

When normalized correctly, it will be between -1 and 1. It is 1 when $f=g$ and $l=0$, and it is -1 when $f=-g$ and $l=0$. Here is the complete cross correlation formula:

$$(1.4) \quad \phi_{f,g}(l,k) = \frac{\sum_{i=k-N/2}^{i=k+N/2} f(i)g(i+l)}{\left[\sum_{i=k-N/2}^{i=k+N/2} f^2(i) \sum_{i=k-N/2}^{i=k+N/2} g^2(i+l) \right]^{1/2}}$$

In the frequency domain:

If $\Phi_{f,g}(\omega)$ is the fourier transform of convolution, we have:

$$(1.5) \quad \Phi_{f,g}(\omega) = F^*(\omega)G(\omega)$$

which has lots of phase and DC problems.

Semblance with Lag

Cross correlation is very sensitive to phase differences. A better choice is Semblance, a measure of the coherence between traces. This is not usually computed with lag, but we need that here:

$$(2.1) \quad S(l,k) = \frac{\sum_{i=k-N/2}^{i=k+N/2} (f_i + g_{i+l})^2}{2 \sum_{i=k-N/2}^{i=k+N/2} (f_i^2 + g_{i+l}^2)}$$

This is approximately the energy of f and G divided by the mean energy of each component. It is between 0 and 1, and is 1 when f=g and l=0.

In our implementation, we assume that the g_{i+l} is an event, i.e. f and g have different lengths, and we are interested in the results of passing g over f . Let N_f and N_g be the lengths of f and g respectively. In our case,

$$(2.2) \quad N_g \leq N_f$$

Then in equation (2.1) f and g are extended with 0 values as necessary, and we have

$$(2.3) \quad S(l) = \frac{\sum_{\substack{i=l \\ 0 \leq i < N_f}}^{N_g+l} (f_i + g_{i-l})^2}{2 \sum_{\substack{i=l \\ 0 \leq i < N_f}}^{N_g+l} (f_i^2 + g_{i-l}^2)}, \quad l = -N_g, \dots, (N_f - 1)$$

that is we sum whenever f and g overlap, that is in the region where $0 \leq i - l < N_g$ and $0 \leq i < N_f$.

Remember that l is the left hand endpoint of g and so one should add $N_g / 2$ to l to find the lag corresponding to the center of the event.

It turns out that when the length of f and g are very different, the semblance for where they partially overlap can be biased. So, instead of (2.3) we use only those points for which f and g completely overlap, making use of the assumption (2.2).

$$(2.4) \quad S(l) = \frac{\sum_{\substack{i=l \\ 0 \leq i < N_f}}^{N_g+l} (f_i + g_{i-l})^2}{2 \sum_{\substack{i=l \\ 0 \leq i < N_f}}^{N_g+l} (f_i^2 + g_{i-l}^2)}, \quad l = 0, \dots, (N_f - N_g - 1)$$

There is no clear frequency domain analogue. If $|f| \gg |g|$, then $S(l) \sim 1/2$. If $f \sim g$ then $S(l) \sim 1$. If $f \sim -g$ then $S(l) \sim 0$. In general, it can be characterized as the energy of the sum divided by the mean energy of the components of the sum. This method is more robust with respect to phase distortions.

For more than two samples, (2.1) can be generalized:

$$(2.5) \quad S(k) = \frac{\sum_{i=k-N/2}^{i=k+N/2} \left[\sum_{j=1}^M f_{ij} \right]^2}{M \sum_{i=k-N/2}^{i=k+N/2} \sum_{j=1}^M (f_{ij}^2)}$$

Writing the equation this way makes it clearer that we are comparing energy of the sum to the mean energy of the components. However, for notational convenience the lag, which is different for each j , is applied the f_{ij} before summing.

In any case, we use equation (2.4) in RunSemblance and in [Semblance](#). To make the array reference the center of the event, we pad $N_g / 2$ zeros on the beginning and the end of the resulting array. Then the returned floating point array is numbered $[0, N_f)$, and the i^{th} point represents the value of the semblance when the middle of the event is lined up with the i^{th} point.